Autonomous Vocal and Backing Track Mixing

Master Project Proposal

Kelian (Mike) Li

Music Informatics Group

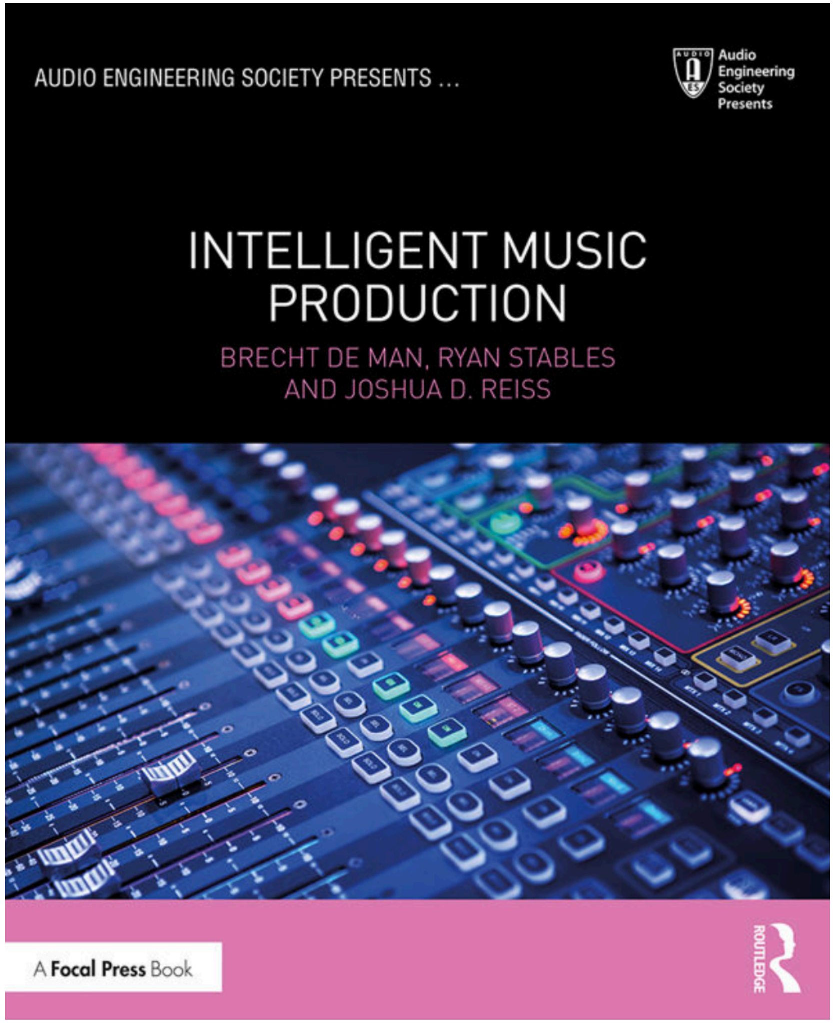
Georgia Tech Center for Music Technology
College of Design

Motivation

Over 10 years of research on Multitrack mixing

No significant achievements yet

How about an easier task?



AUDIO ENGINEERING SOCIETY PRESENTS …

Audio Engineering Society Presents

INTELLIGENT MUSIC PRODUCTION

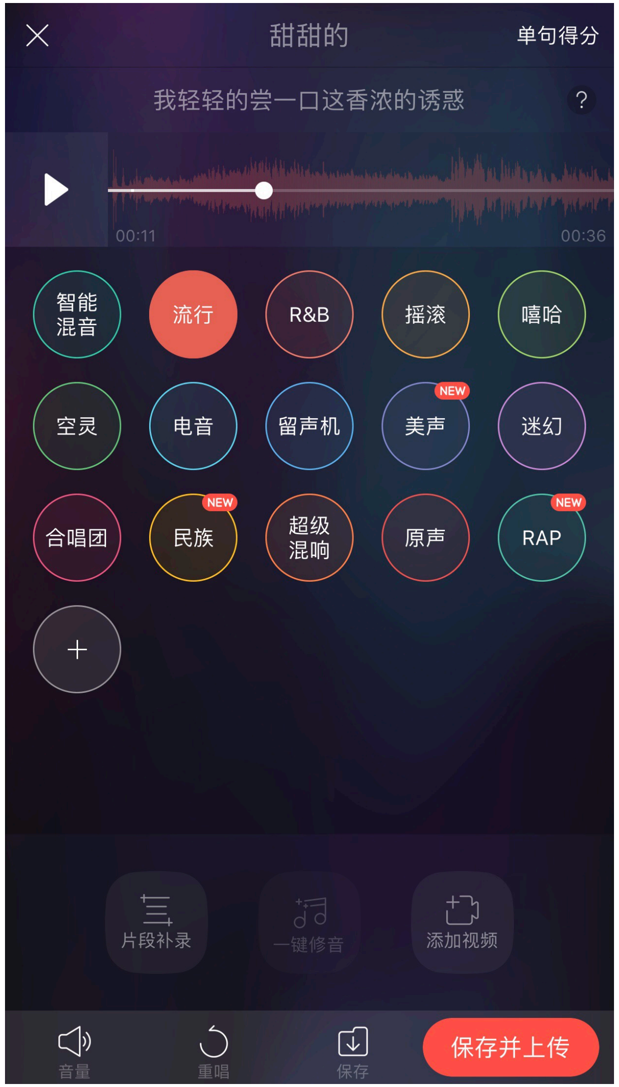
BRECHT DE MAN, RYAN STABLES AND JOSHUA D. REISS

A Focal Press Book

ROUTLEDGE

Motivation

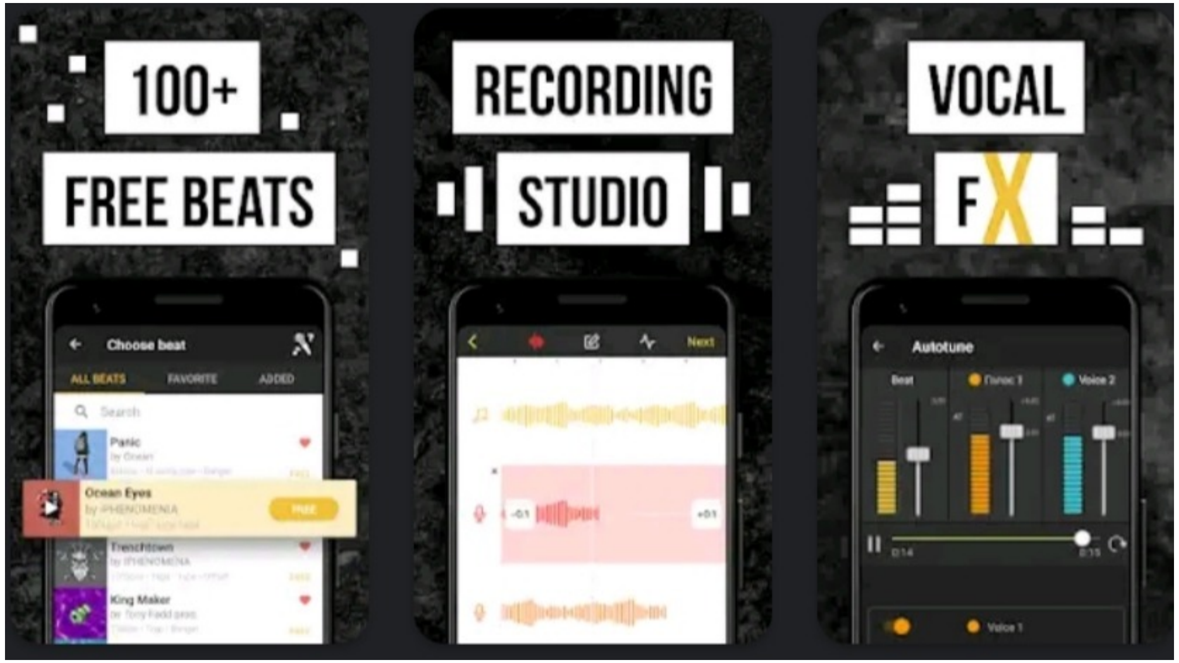
- Karaoke apps



我轻轻的尝一口这香浓的诱惑

Motivation

- Karaoke apps
- Amateur music makers



Related Work

- Knowledge-based:
  - Mixing rules and mix analysis

Related Work

- Knowledge-based:
  - Mixing rules and mix analysis
- Data-driven:
  - The main challenge is the data collection of mixing parameters

Related Work

- Knowledge-based:
  - Mixing rules and mix analysis
- Data-driven:
  - Extract mixing parameters from **paired raw tracks and human-mixed** tracks
    - Reverse engineering of a mix[1]
    - Differentiable signal processing chain[2]
    - Gradient approximation on black-box audio effects[3]

Related Work

- Knowledge-based:
  - Mixing rules and mix analysis
- Data-driven:
  - Extract mixing parameters from **paired raw tracks and human-mixed** tracks
    - Reverse engineering of a mix[1]
    - Differentiable signal processing chain[2]
    - Gradient approximation on black-box audio effects[3]
  - End-to-end audio transformation[4]
    - No further control by users

Related Work

- Knowledge-based:
  - Mixing rules and mix analysis
- Data-driven:
  - Extract mixing parameters from **paired raw tracks and human-mixed** tracks
    - Reverse engineering of a mix[1]
    - Differentiable signal processing chain[2]
    - Gradient approximation on black-box audio effects[3]
  - End-to-end audio transformation[4]
    - No further control by users

Lack of Data

Georgia Tech | Center for Music Technology
College of Design

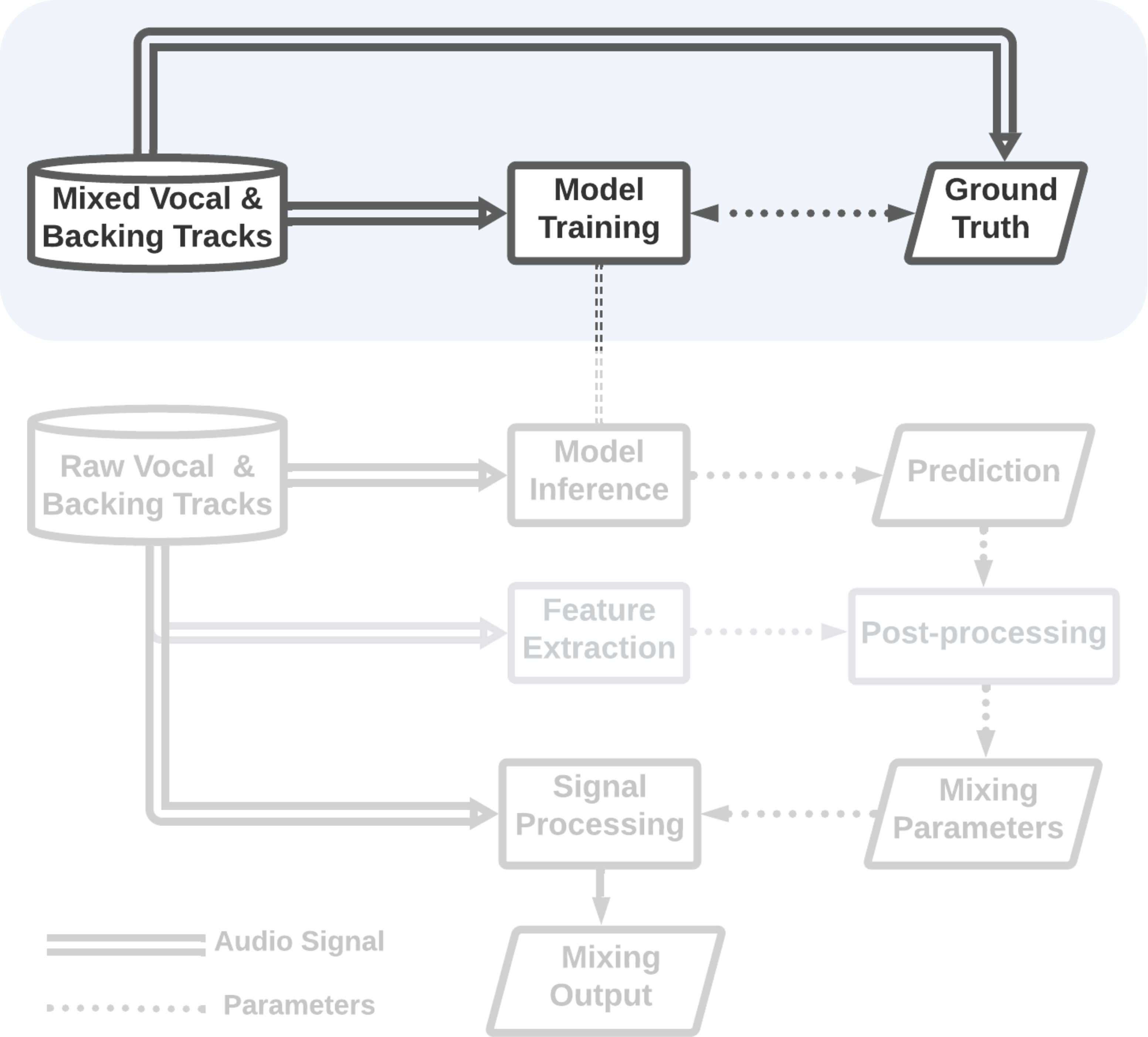
Baseline System

- Level balance
  - −3 dB vocal-to-mix ratio
- Compression
  - 14 dB loudness range
- EQ
  - Frequency masking
- Reverb
  - Linear mapping from tempo to reverb time

Proposed Method

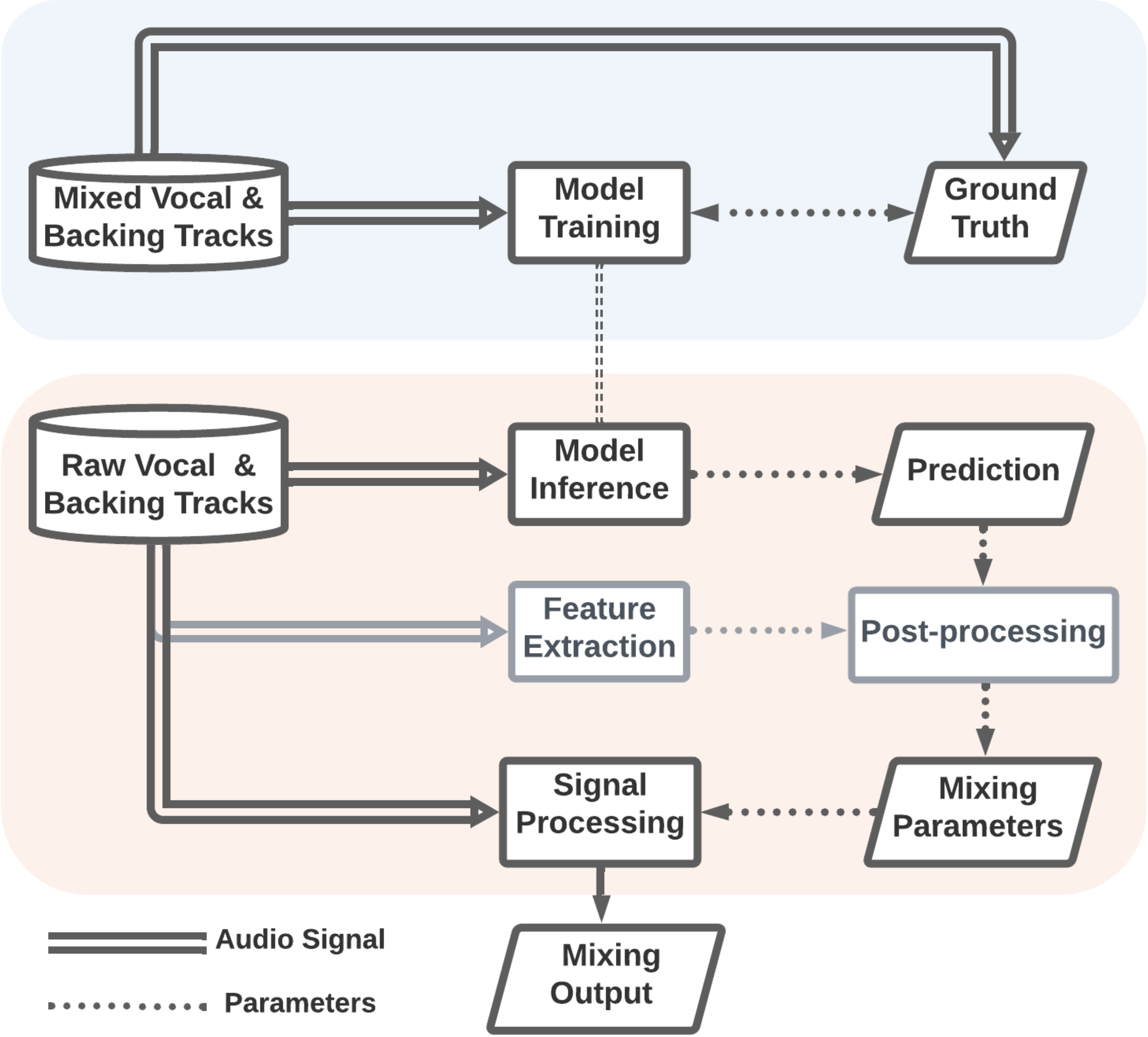
- Data-driven
- Maps the input audio to mixing parameters
  - Outputs mixing parameters which allows human adjustment
- Requires only mixed vocal and backing tracks for training
  - Raw vocal tracks are not needed

Model Training



Ground truth is the direct or intermediate mixing parameters

Model Training



Ground truth is the direct or intermediate mixing parameters

Model Inference

Optionally, convert intermediate model outputs into direct mixing parameters

Proposed Method
-Level Balance and Compression

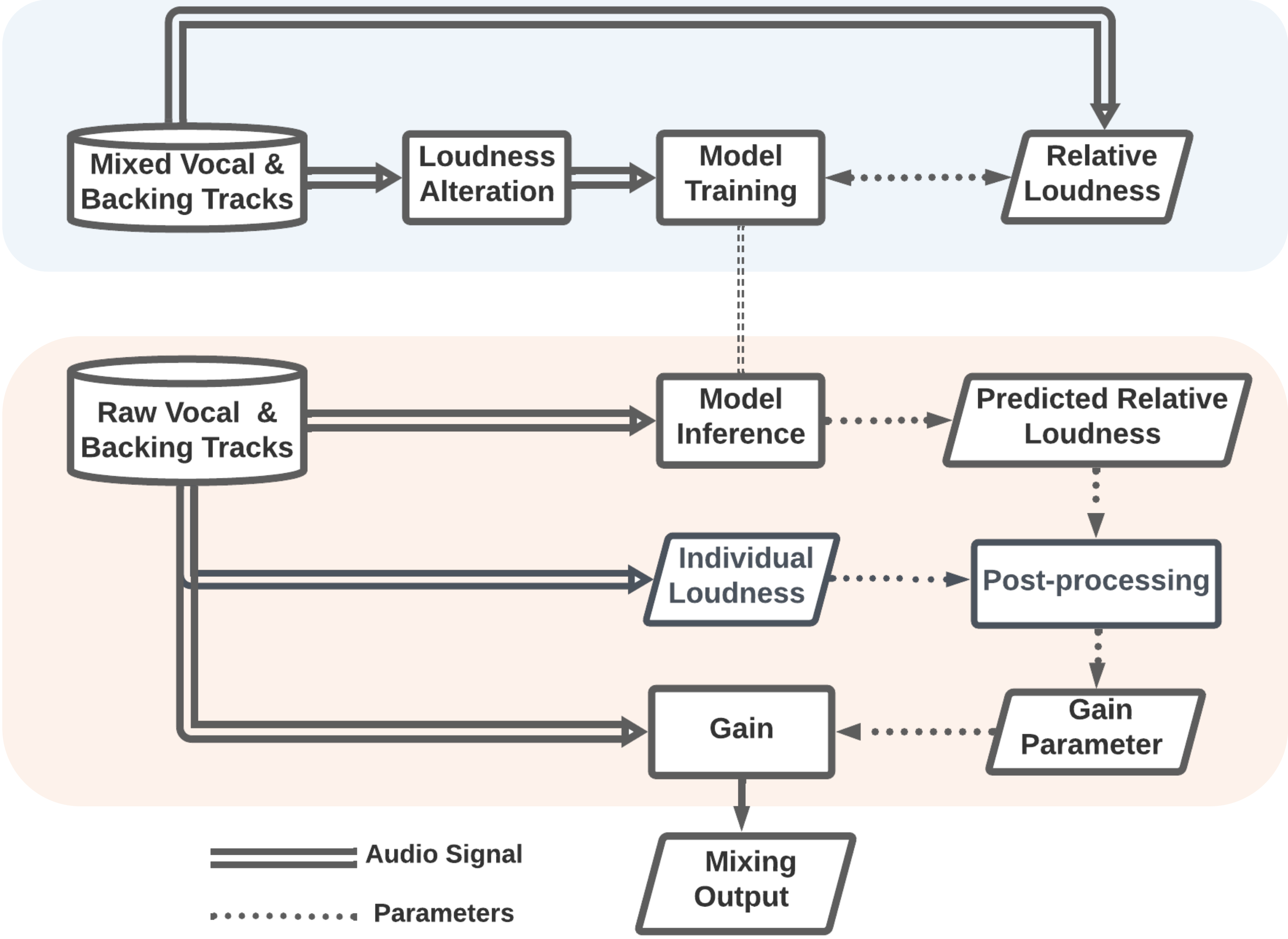
- The model outputs intermediate audio features (relative loudness and loudness range)

- Post-processing converts the intermediate features into mixing parameters

Proposed Method
-Level Balance

Model Training

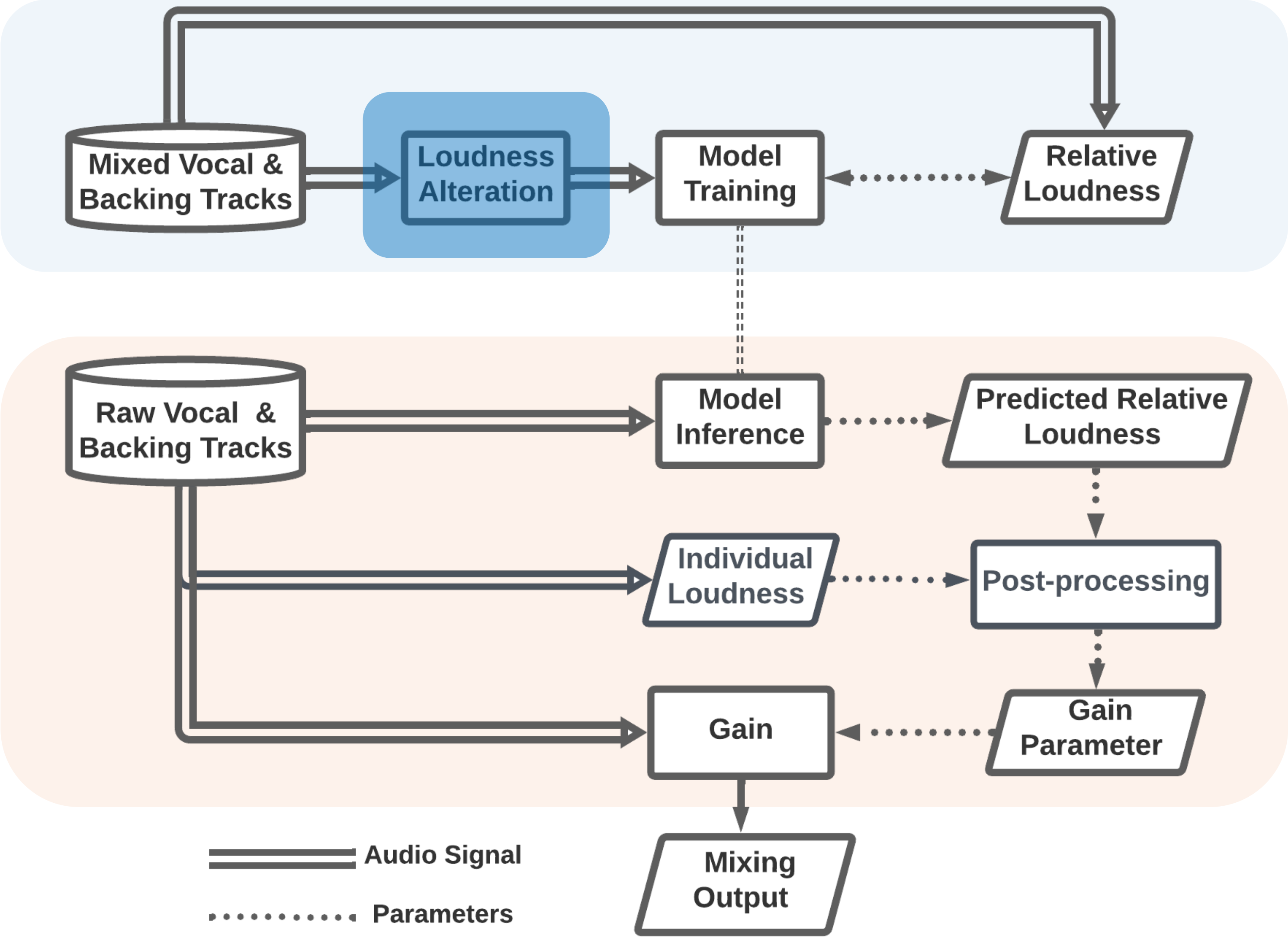
Model Inference



Proposed Method
-Level Balance

Model Training

Model Inference

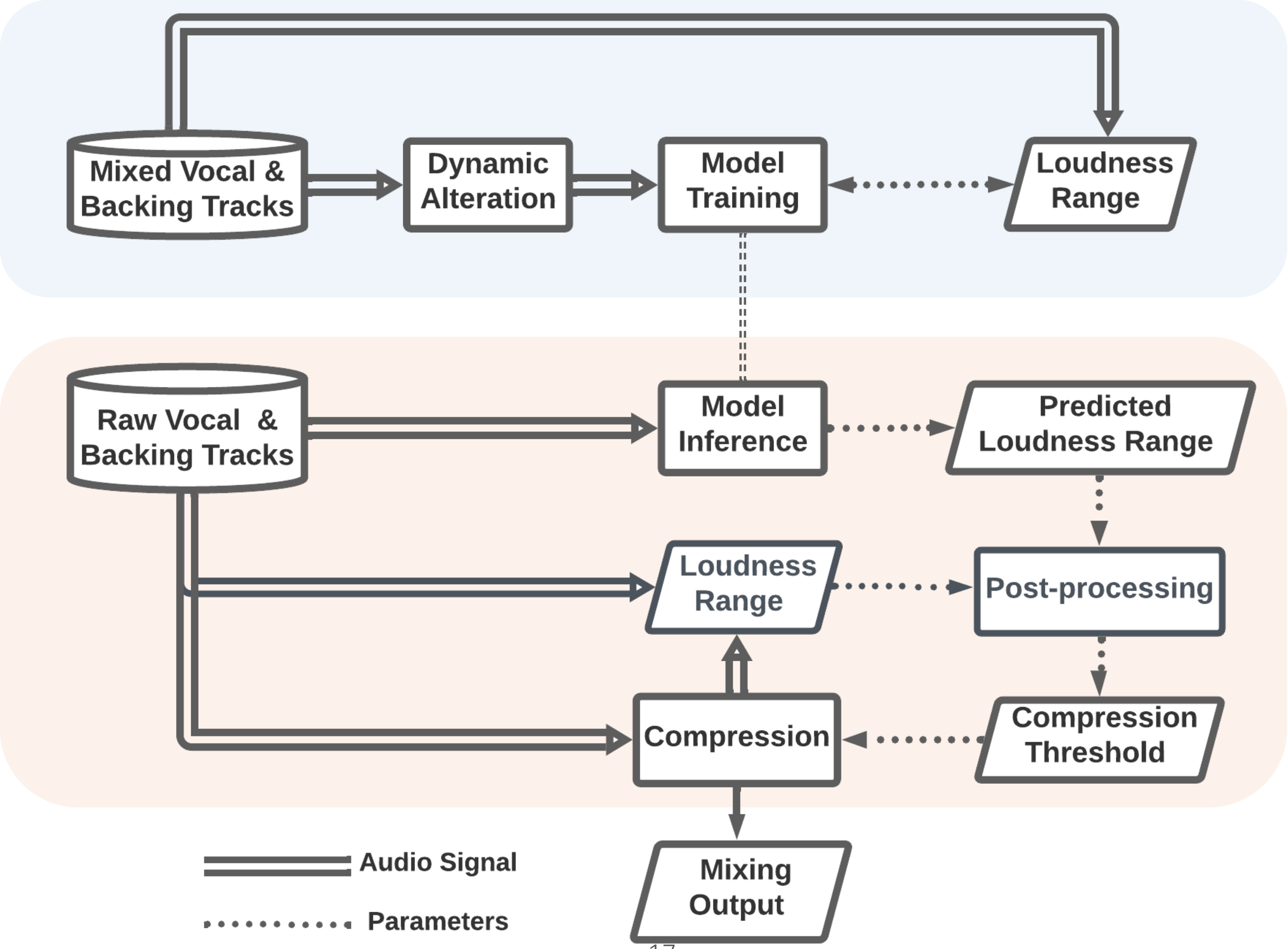


The model should learn to mix, instead of extracting parameters directly

Proposed Method
-Compression

Model Training

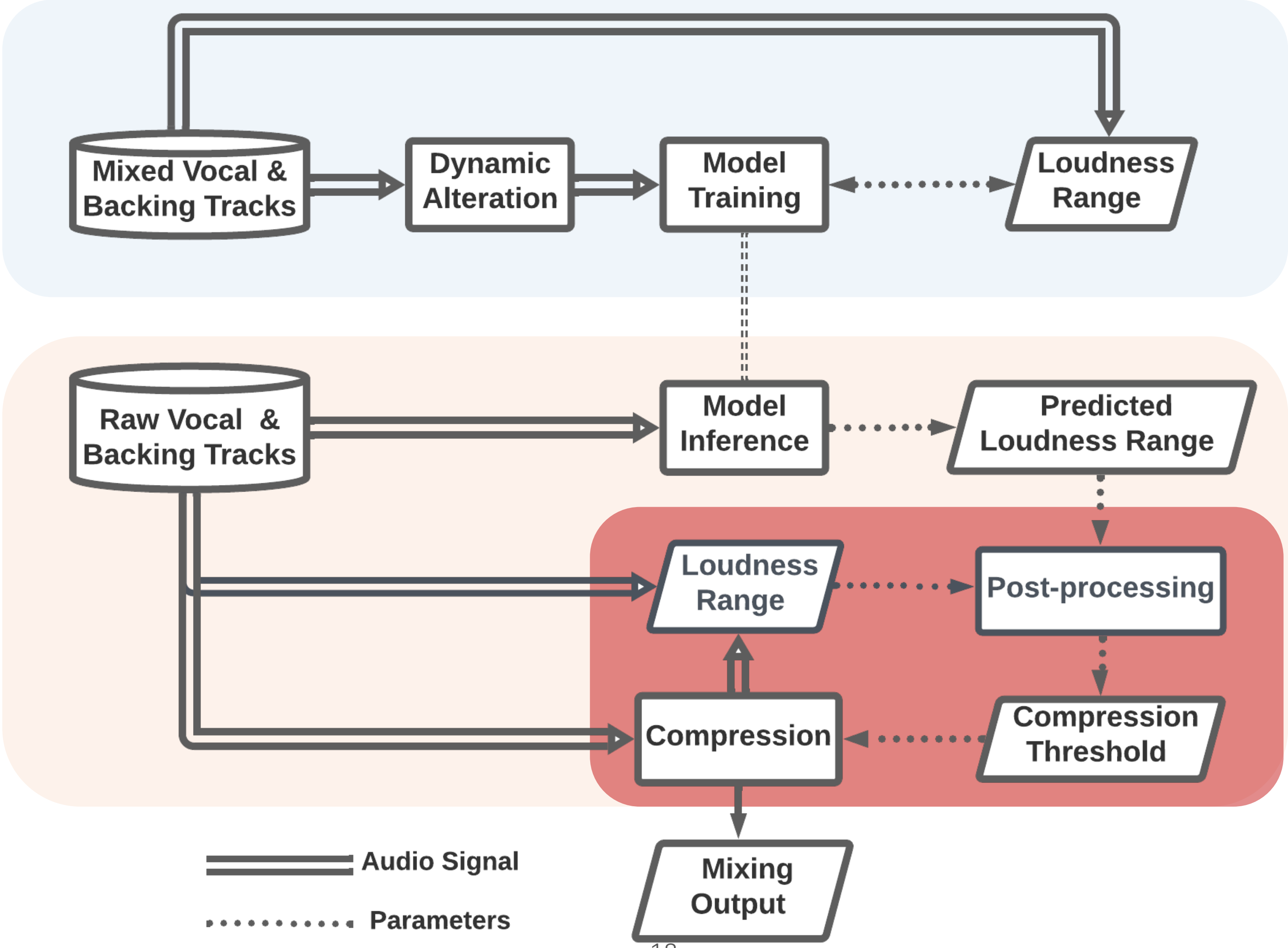
Model Inference



Proposed Method
-Compression

Model Training

Model Inference



An iterative process to find the compression threshold of the targeted loudness range

Georgia Tech Center for Music Technology College of Design

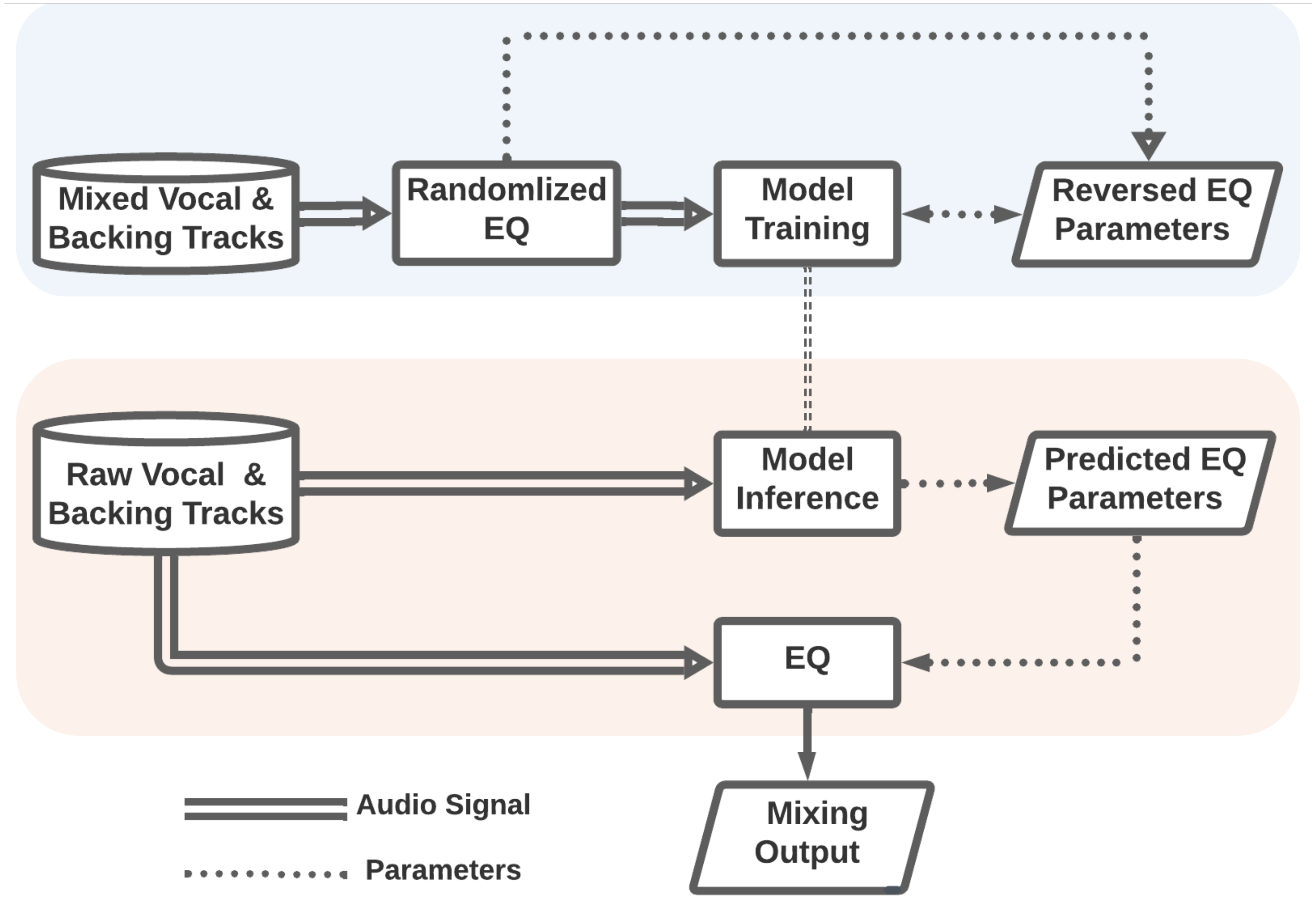
Proposed Method
-Equalization

- The “raw” tracks for training are self-generated by applying EQ to the mixed vocal tracks. The corrected parameters are known.

Proposed Method
-Equalization

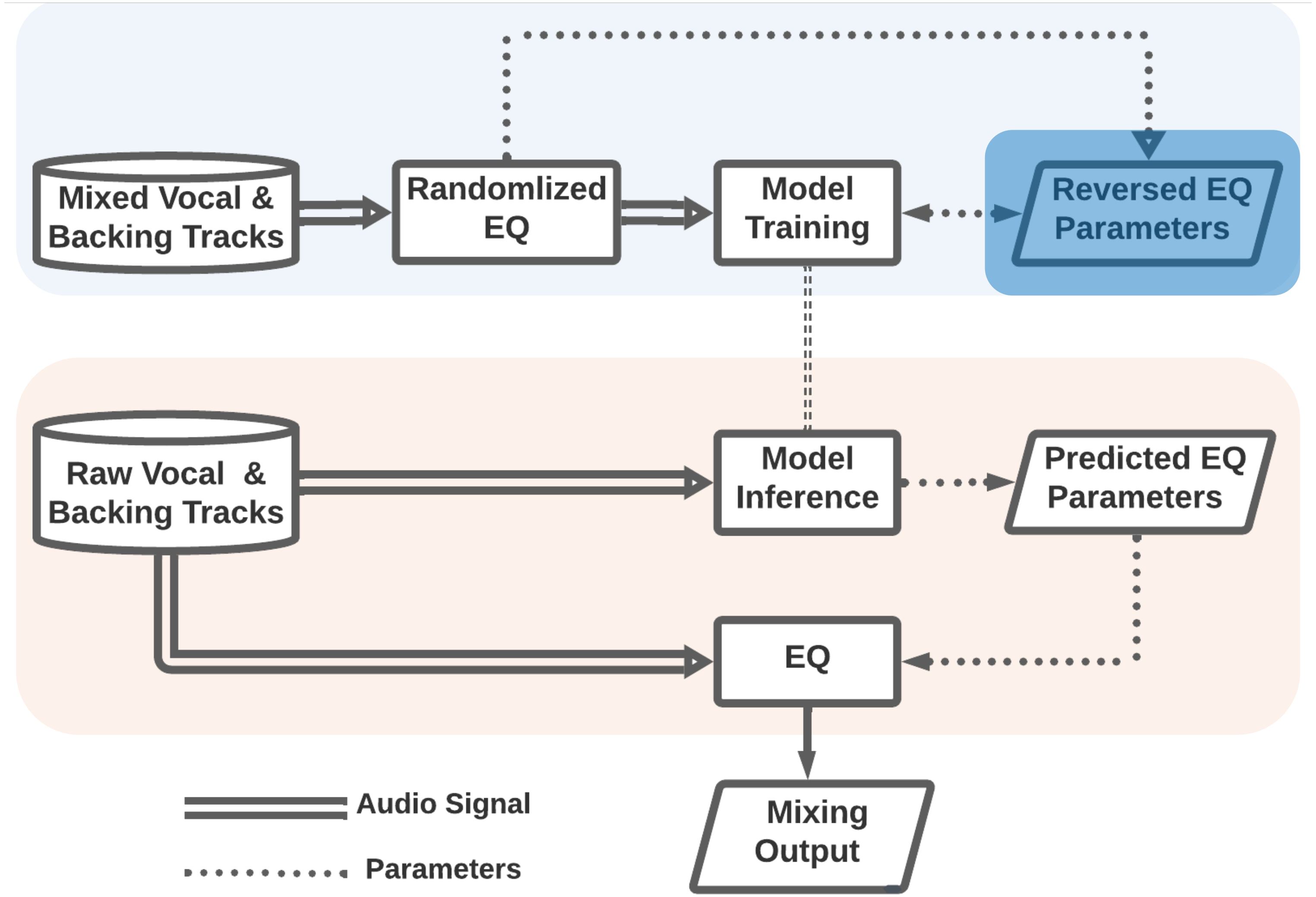
Model Training

Model Inference



Georgia Tech Center for Music Technology
College of Design

Proposed Method -Equalization



If the mixed vocal is boosted at some center frequency, we should learn to cut at that frequency.

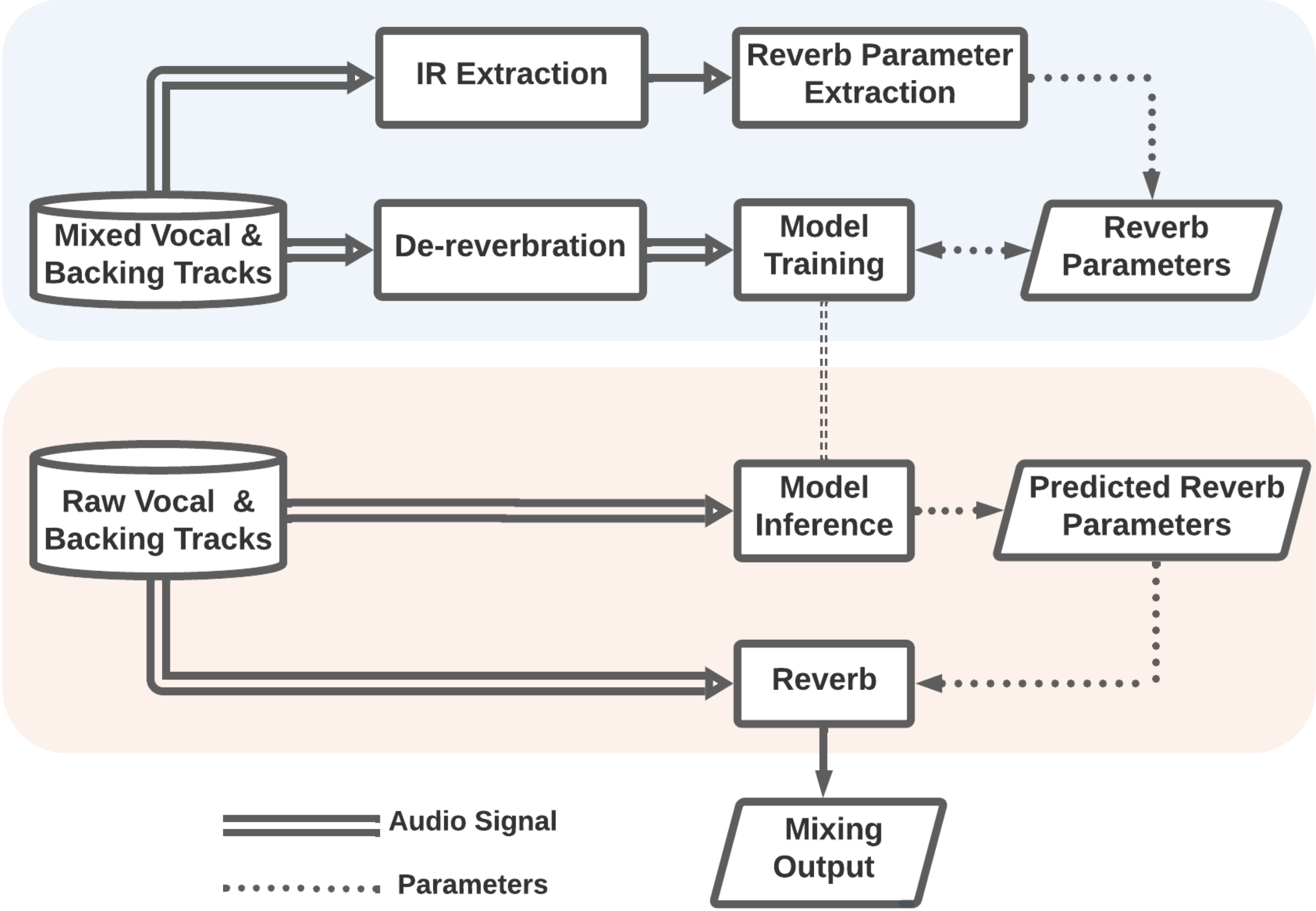
Proposed Method
-Reverberation

- Extracts the reverb impulse responses by a commercial plugin
- Uses genetic optimization to approximate the reverb parameters for the impulse responses

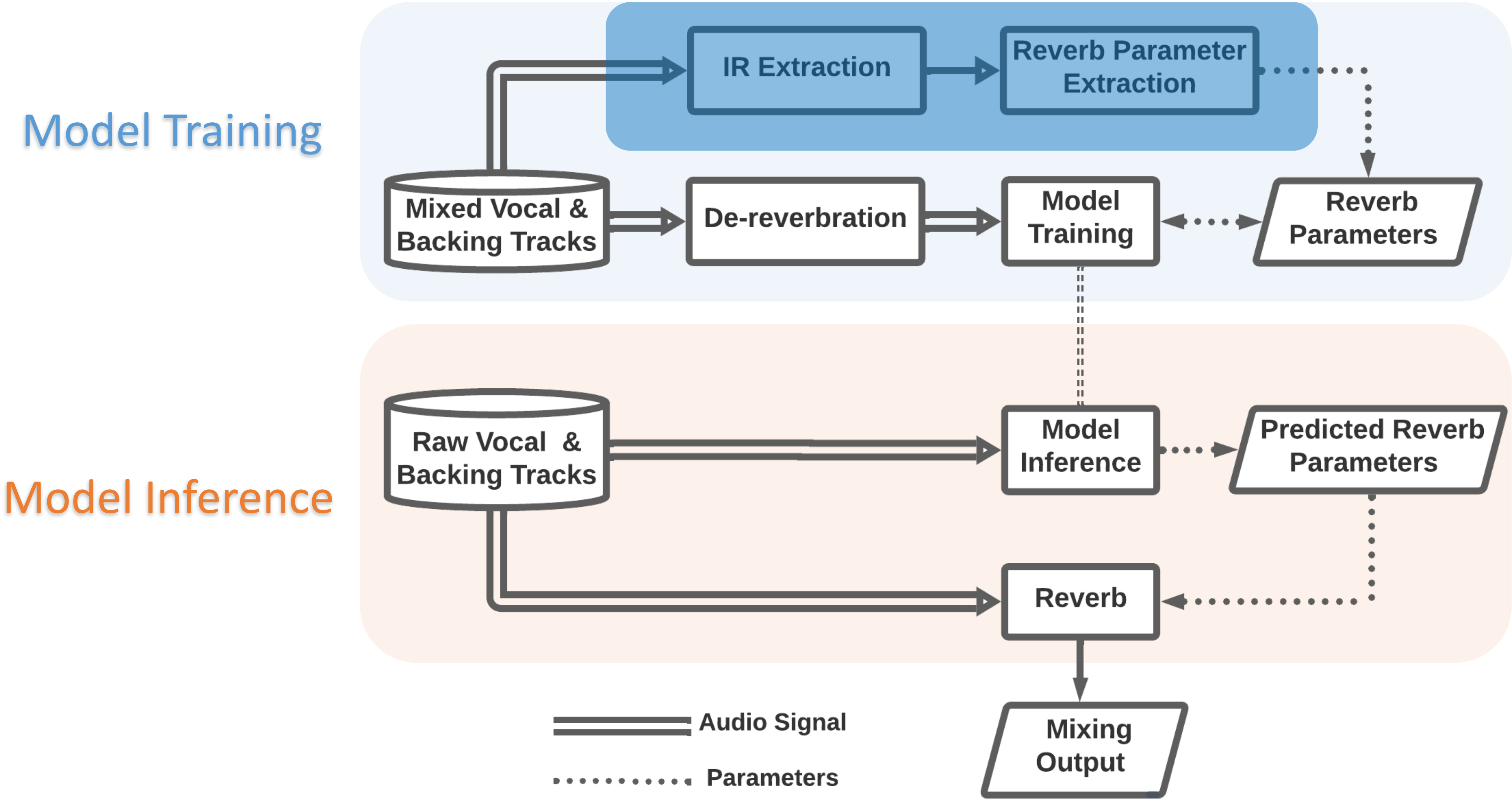
Proposed Method
-Reverberation

Model Training

Model Inference

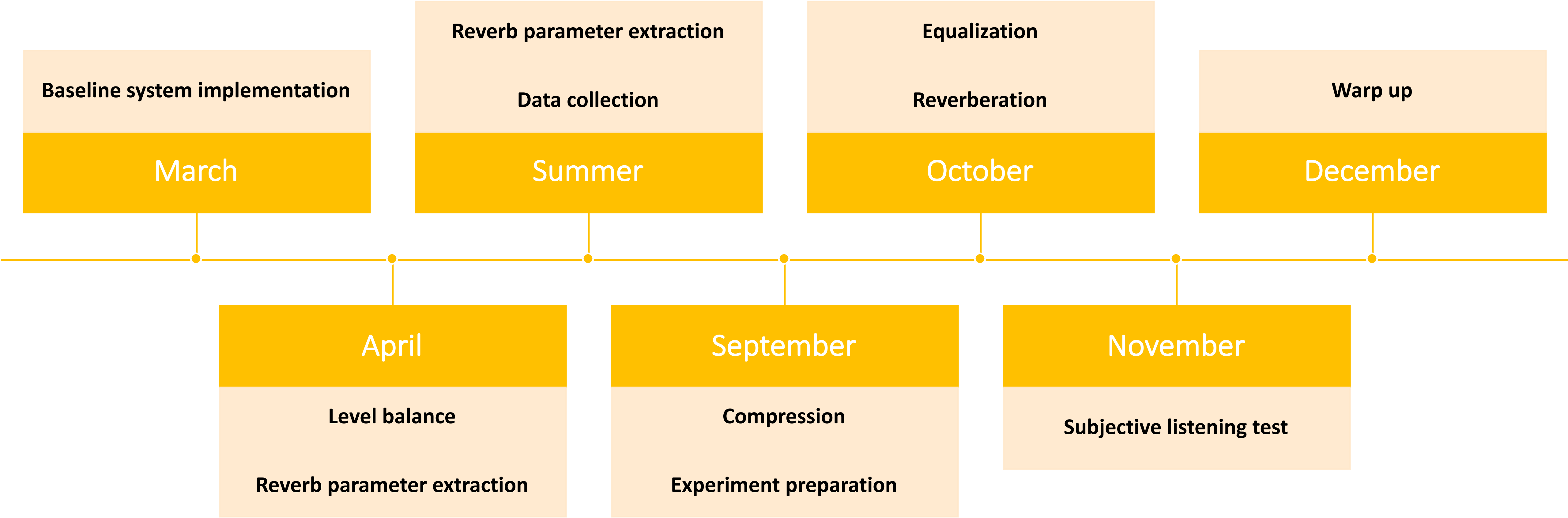


Proposed Method -Reverberation



Georgia Tech Center for Music Technology
College of Design

Timeline



March: Baseline system implementation

April: Level balance; Reverb parameter extraction

Summer: Reverb parameter extraction; Data collection

September: Compression; Experiment preparation

October: Equalization; Reverberation

November: Subjective listening test

December: Warp up

References

[1] D. Barchiesi and J. Reiss, “Reverse engineering of a mix,” *Journal of Audio Engineering Society*, vol. 58, no. 7/8, 2010.

[2] C. J. Steinmetz, J. Pons, S. Pascual, and J. Serra, “Automatic multitrack mixing with a differentiable mixing console of neural audio effects,” in *ICASSP*, IEEE, 2021.

[3] M. A. Martinez Ramirez, O. Wang, P. Smaragdis, and N. J. Bryan, “Differentiable signal processing with black-box audio effects,” in *ICASSP*, IEEE, 2021.

[4] M. Martinez Ramirez, D. Stoller, and D. Moffat, “A deep learning approach to intelligent drum mixing with the wave-u-net,” *Journal of Audio Engineering Society*, vol. 69, no. 3, 2021.

Thank you!

Georgia Tech Center for Music Technology
College of Design